Inconsistencies in Hausa Digraphs: 
Implications for Corpus Planning

Clifford Irikefe Gbeyonron

Abstract
Hausa is a well-developed language. Formal and non-formal corpus planning has engendered the availability of Hausa orthography. Thus, Hausa alphabets in 'ajami' and 'boko' forms abound. Of interest in this study is the digraph of the 'boko' form of written Hausa and the consistencies therein. Relying on the theory postulated by Williamson (1984), it was found that the digraphs 'sh' and 'ts' are inconsistent with the other digraphs in the language when considered from their phonological realizations. In consequence, it is recommended that the 'monographic' form of 'sh' and 'ts' as found in 'ajami' should be adopted. In order to achieve this, some plausible alphabets that could represent 'sh' and 'ts' are presented. These could be considered by corpus planners for adoption or adaptation.

Introductory Background
According to Bartsch (1987:278) corpus planning “… comprises language elaboration and cultivation.” As such, one of the primary concerns of corpus planning is graphization (Junaidu, 2010; Karam, 1974). This involves the development of orthographies, script replacement and script reform. Hausa language has undergone both status and corpus planning processes in consequence it is a well-developed language. This can be appreciated from the fact it had had a long history of codification as far as written Hausa is concerned. Thus, it has an orthography which has experienced replacement – even though the hitherto orthography is used pari pasu the one that replaced it – and script reform. This notwithstanding, there are some inconsistencies in the 'digraph' of Hausa thus necessitating further orthographic reforms in respect of the language. It is in the light of this that this study explores the inconsistencies that abound the Hausa 'digraphs'.

The Problem
Some scholars in Hausa linguistics have identified inconsistencies in the orthography of Hausa (Yalwa, 2013). While some of the inconsistencies are cross-linguistic (they pertain to cross-national varieties of Hausa orthography),
some have been intra-linguistic (in respect of divergent views of scholars on some national variety). However, there is a paucity of literature in respect of the inconsistencies that exist in the Roman script 'digraph' of Hausa. As a result, this study was conceived to dissuade the use of some terms for Hausa consonant clusters and consonant digraphs and also postulate new ways of presenting Hausa digraphs in the Roman script.

**Objective**

The primary objective of this study is to investigate the inconsistencies that exist in the 'digraphs' of Hausa Roman script (Boko) with the intention of proffering positions for orthographic reform. In order to achieve this, the following questions are generated to serve as guides:

- How many digraphs does Hausa Roman script have?
- Do the Hausa Roman script digraphs provided by some scholars have the same characteristics?
- How can corpus planning check the inconsistencies that abound so as to engender orthographic reform?

**Orthography and Orthographic Reform: An Overview**

According to Williamson (1984:7); “Orthography consists of the symbols and the rules that are followed in writing a language.” Dobrovolsky (2011:538) defines orthography as: “The set of conventions for representing language in written form …” In his own submission, Finegan (2012:539) states that orthography is: “A system of spelling used to achieve a match between the sound system of a language and the alphabet representing it.” This implies that the orthography of a language is the norm that guides the graphic or visual symbols that would be used to represent the spoken form of a language. In order to achieve the aim of orthography or written language, linguists have utilized logographic writing (use of logograms) and phonographic writing (use of syllabary, consonantal alphabets and alphabets (Dobrovolsky, 2011, Fromkin, Rodman & Hyams, 2011). Preceding these forms of writing were pictograms, ideograms, cuneiforms, Rebuses and hieroglyphics (Denham & Lobeck, 2013). Most Nigerian languages use both the Roman and Arabic scripts. It should be noted that Igbo language had a pre-Roman and Arabic orthography. For instance, Achebe (2012:192) reports the existence of Nsibidi, an ancient writing invented by the Ejaghams and adopted by the Igbos, Efiks, Anangs and Ibibios and asserts that the “… very existence of this alphabet dating back to the 1700s without any Latin and Arabic antecedent, is a rebuke to all those who have claimed over the centuries that Africa has no history, no writing and no civilization!”
Despite the fact that languages all around the world can adopt or adapt the Greek, Latin, Roman, Arabic or other non-European and Arabic scripts (Dobrovolsky, 2011); Fromkin et al (2011:540) state that: “Even in today's world, many spoken languages lack a writing system …” Thus, this situation has prompted linguists in Nigeria in collaboration with the Nigerian Educational Research and Development Council (NERDC) to underscore the development of orthographies for languages that do not have. The Linguistic Association of Nigeria (LAN) has also promoted the development of new orthographies and encouraged advocacy for the reform of existing ones.

A systematic process is followed for one to arrive at the orthography of a particular language. Thus, a team of corpus planners must come together to develop and validate the orthography based on some norms so as to achieve the aim of graphization and codification. As a result, Halliday 1977; 2003:558) reports: “Many languages have been the subjects of spelling reforms in the past hundred years, including Dutch, Norwegian, and Russian.” In the case of Hausa, Yalwa (2013:600) reports that The Literature Bureau embarked on orthographic reform in 1933 for the purpose of “… finding new ways of writing Hausa in order to minimize the influence of English spelling on Hausa that characterized the Vischer's spelling (such as the usage of 'c' instead of English 'ch', introduction of the hooked letters etc).” Since then, series of harmonization processes have been embarked upon with the intention of ensuring reforms in the orthography of Hausa (Egbokhare, 2011). Suffice it to say that Williamson (1984:10) cautions in respect of orthographic reform stating: “If we try to change too many things, people will just refuse to accept them and the orthography will be controversial.” The implication of this counsel in respect of corpus planning, and the development of orthography in particular, is what so ever writing system to be developed, adopted or adapted should be based on empirical and meticulous exercise.

**Hausa Alphabets and the Presentation of 'Digraphs': The Inconsistencies Within**

One of the techniques used in phonographic writing is the alphabetic system. Denham and Lobeck (2013:444) submit: “In alphabetic systems, each symbol represents a sound.” Hausa appears to be using the alphabetic system in both Roman and Arabic scripts. Amfani (2007:132) asserts: “The present Hausa orthography is almost fully phonetic. That is to say, the orthography or the writing system fully captures the way and manner sounds and word are uttered.” Newman (2000:726) reports: “HAUSA has two writing systems. One of them, termed boko (<English 'book'?), is based on the Roman alphabet … The other, termed ajami, is based on the Arabic alphabet.” An online material obtained from www.aflang.hummet.ucla.edu/Hausa/Pronunciation/writing.html provides: “Learned people have written Hausa in ajami since at least the early
19th century when it was pressed into use primarily to write Islamic poetry in praise of Allah and the Prophet and to expound Islamic doctrine.” Thus prior to the annexation of Northern Nigeria by the British in 1903, Hausa has had a writing system it borrowed from Arabic. In the case of boko, Yalwa (2013:599) reports: “Hausa was first reduced to writing using the BOKO (Roman script) in the 1840s.” However, he underscores that Hanns Vischer's document of 1912 entitled *Rules for Hausa Spelling* is the first work that provided direction in respect of formal Hausa orthography. Of particular interest in this work is how Hausa 'digraphs' are presented in boko (Roman script) and the inconsistencies that abound.

The numerical strength of Hausa alphabets have been presented differently by different scholars of Hausa linguistics. For instance, Galadanci (1976, 2003:108) states: “The total number of Hausa consonants disd thirty-one (31).” He presented the consonants in the upper and lower cases. The ones in lower case are presented as: (i) b, c, d, f, g, h, j, k, l, m, n, r, s, sh, t, ts, w, y, z. (ii) ɓ, ɗ, ƙ, ’y. (iii) ky, gy, fy, ƙy, kw, gw, ƙw. The alphabets in (i) are normal Roman alphabet; those in (ii) are hooked or apostrophe carrying alphabets while the ones in (iii) are clusters. Note that the digraphs 'sh' and 'ts' are grouped in the normal Roman alphabets category. Centre for the Study of Nigerian Languages (1977:ix) in the guide to the use of the dictionary *Modern Hausa-English Dictionary* provides that all main entries “… are arranged in alphabetic order according to the standard Hausa alphabet: ', a, b, ɓ, c, d, d, e, f, g, h, i, j, k, k, l, m, n, o, r, s, t, u, w, y, ’y, z.” This submission reveals that only the normal consonants, the hooked consonants and the apostrophe carrying consonants are presented explicitly. The digraphs 'sh' and 'ts' are treated under 's' and 't' respectively. In the case of the consonant clusters, they appear under their cognate 'head' normal or hooked consonants. Jinju (1980) provides the following as the consonant alphabets of Hausa: B, B, C, D, D, F, FY, G, GW, GY, H, K, KW, KY, K, KW, KY, L, M, N, R, S, SH, T, TS, W, Y, ’Y, Z. This presentation is different from ones presented by Galadanci and Centre for the Study of Nigerian Languages. It can be realized from this presentation that the normal, hooked and apostrophe carrying consonants in addition to digraphs and clusters are lumped together. The implication of this is one may think that the digraphs and the clusters have the same characteristics. This means that one may think that the clusters are digraphs or the digraphs are also clusters. Sani (2009) presented the Hausa consonants in slanted lines. They are: /b, ɓ, m, f, fy, t, d, d, l, r, s, n, z, ts, c, j, sh, n, y, k, k, g, ky, ky, gy, kw, kw, gw, w, ’, h, ’y/. The approach here is distinct from the ones presented earlier. Even though the normal, hooked and apostrophe carrying consonants in addition to digraphs and clusters are lumped together, they are presented in slanted lines (which in some cases are inconsistent with their phonological alphabetic realizations) and grouped based on their place or manner of articulation.
The overview presented thus far on Hausa orthographic alphabets agrees with the submission of Yalwa (2013:601) who reviewed a lot of works written in respect of Hausa orthography and thus concludes: “… there are some differences in their usage of the orthographic conventions.” This has generated a lot of criticisms in respect of the incongruence of the positions of scholars on Hausa orthography. Except there is harmonization via orthographic reform, the criticisms and counter criticism will rage on.

Methodology and Theoretical Framework

Agbo (2013) identifies two approaches to the study of orthographies. They are the priori and the posteriori approaches. While the priori approach attempts to conduct a pilot study on an emerging orthography of a language, the posteriori approach investigates the merits and the demerits of an established orthography with the intention of recommending it for adoption, adaptation or total reform. In the light of this, this study adopts the posteriori research design to investigate the merit and the demerit of Hausa 'digraphs'. In order to achieve this, existing corpora were analyzed in respect of the usage of the alphabets. In addition, reference grammars were also reviewed to establish congruence or incongruence in the presentation and description of Hausa 'digraphs' and 'clusters’.

This study utilizes Williamson (1984) principle of accuracy, consistency, convenience, harmonization and familiarity as theoretical framework. In particular, the principle of consistency was adapted for data analysis while the principle of harmonization was adopted for the purpose of recommendations for orthographic reforms. According to Williamson (1984:8); “We say an orthography is consistent if the same sound or word is always written in the same way anywhere it appears.” Relying on the principle of alphabetic system where each symbol is expected to represent a sound, Williamson's principle of consistency is adapted by this study thus: “Symbols should be classified as the same when they behave in the same way in different environment.” Furthermore, Williamson (1984:10) postulates in the principle of harmonization that if a letter is used in one language “… to represent a certain sound, every effort should be made to avoid using that same letter to represent a completely different sound in another language.” This principle provides the basis for the recommendations made by this study in respect of orthographic reforms.

Results

The following are the major findings of this study. The findings are presented as answers to the research questions raised by this study.
How many digraphs doe Hausa Roman script have?

Relying on the definition of Akmajian, Demers, Farmer and Harshish (2001:576) which provides that a digraph is: “A sequence of two letters used to spell a single sound”, it was found that Hausa has two digraphs viz. 'sh' and 'ts'. Note that Newman (1997:xiv) considers ‘y as a digraph where she states: “The digraphs ts and ‘y represent glottalized s and y respectively.” This work did not present ‘y as one of the digraphs of focus because it is represented as a monograph in the Hausa orthography used in Niger Republic.

Do the Hausa Roman script digraphs provided by some scholars have the same characteristics?

When the researcher considers the definition provided by Akmajian et al (2001) above in consonance with the position of Fromkin et al (2011:553) which states: “Some languages use two letters together – called a digraph – to represent a single sound”, it was found that what some scholars consider as digraphs do not have the same characteristics. For example, Newman (2000:727) states: “The digraph ts, which represents an ejective sibilant … Like ts, the digraph sh … The digraphs involving w and y e.g. gw, gy, fy, kw, ky, etc are typically treated as sequences of letters for orthographic purposes, notwithstanding their phonemic status.” Similarly, Bunza (2004) considers 'kw', 'gw' and 'gy' on one hand, and 'sh' and 'ts' as tagwayen bakake (clusters/digraphs). Given that sh and ts represent a single sound in each case and the so called digraphs involving w and y represent two sounds in each case (clusters), it is concluded that what some scholars consider as digraphs are not digraphs but clusters. This is because Amfani (2007:131) asserts that fy, gw, gy, kw, kw, ky and ky “… are produced through double articulation.”

How can corpus planning check the inconsistencies that abound so as to engender orthographic reform?

Findings from secondary research reveals that Hausa has undergone several orthographic reforms (cf. Egbokhare, 2011; Newman, 2000; Yalwa, 2013). As a result, it is concluded that further orthographic reforms can check the inconsistencies identified by this study as postulated in the sections of this work.

Discussions and Implications for Corpus Planning

The findings of this study reveals that what some scholars consider as digraphs are consonant clusters. Thus inconsistencies exist when consonant clusters and real digraphs are categorized as digraphs. This has in essence had some effects on the teaching and learning of Hausa where learners are told that both real digraphs and consonant clusters are tagwayen bakake. Reporting on the
evolution of harmonization in respect of Hausa orthography, Yalwa (2013:601) states: “The main change concerns the capitalization of diphthongs and digraphs (in the new decision, only the first letter of the cluster is capitalized).” One can realize from the preceding quotation that the assertion outside the bracket does not contain the word cluster; however the word cluster was mentioned in the assertion in the bracket. Based on the principle of reference, one can conclude that the word cluster as used in the assertion could mean either the word diphthongs or the word digraphs. The writer of this work hypothesizes that the word cluster refers to digraphs in the context. This is because diphthongs were not listed in the chapter from where the quotation was extracted. If, however, it is cluster that is regarded as digraph, Yalwa's position agrees with those of Newman (2000) and Bunza (2004) who lumped both digraphs and consonant clusters and called them digraphs or tagwayen bakake. This is not in agreement with the position of this study. If Newman's and Bunza's positions are allowed to subsist then one can state explicitly that there is inconsistency in the so-called 'digraphs' of Hausa.

Amfani (2007) reports that the Hausa orthography is phonetic in nature. This means that letters capture how sounds are uttered. However, when one analyzes the submission of Denham and Lobeck (2013) where the alphabetic system is characterized by one symbol representing one sound, the real Hausa digraphs are inconsistent with the position of Amfani and, Denham and Lobeck. This is because sh and ts as letters do not in essence represent the sounds the way the other orthographic alphabets and the phonetic alphabets of Hausa do.

The inconsistencies that abound necessitates orthographic reform in respect of the real digraphs of Hausa and by extension their consonant counterparts in the language. It should be noted that ajami pre-dates boko in Hausa. A study of ajami reveals that sh and ts are represented by ش and ظ or ط respectively (Newman, 2000; www.english.ucla.edu/Hausa/Pronunciation/writing.html). Each of these is a single symbol the diacritic notwithstanding. Similarly, Amfani (2007) and Sani (1999) represented sh phonetically as /ʃ/. However, while Amfani (2007) represented ts phonetically as /t/, Sani (1999) represented it as phonetically /s/. Furthermore, Centre for the Study of Nigerian Languages (1977) considers real digraphs and consonant clusters as normal consonants given that they are presented in the entries of the first consonant of either the digraph or the consonant cluster in question. Thus, digraphs and consonant clusters are not listed as separate alphabets in the submission of Centre for the Study of Nigerian Languages (1977). Note that consonant clusters are not listed as separate alphabets in ajami rather users are guided towards the process of joining consonants to form clusters in ajami. Similarly, based on the principle of harmonization, most Nigerian languages and the English language do not treat explicitly physical consonant clusters as separate sequence of alphabets.
Orthographic or spelling reform is not a new thing in linguistics. Fromkin et al (2011:558) reports: “Many languages have been the subjects of spelling reforms in the past hundred years including Dutch, French, Norwegian and Russian.” Amfani (2007:132) traced the history of the development of Hausa orthography in the Roman script to the time of the British colonial rule in Nigeria and adds that after “… Nigeria's independence several conferences were organized to look into the issue of Hausa orthography.” These conferences engendered harmonization and by extension spelling reforms. It is apposite to assert that prior to Nigeria's independence Hausa witnessed orthographic and spelling reforms. For instance, Newman (2000:727) states: “The letter c represents the English sound ch (as in 'church'). In early written works, it was represented as ch as it is in English.” Yalwa (2013:600) agrees and states that in 1933, The Literature Bureau “… took the task of finding new ways of spelling Hausa in order to minimize the influence of English spelling on Hausa that characterized the Vischer's spelling (such as the usage of letter 'c' instead of English 'ch' …).” Thus, if the digraph ch can be changed to c, the digraphs sh and ts can also be represented with one letter in each case instead of the use of a sequence of letters that depicts one sound in each.

**Conclusions and Recommendations**

The findings of this study reveals that Hausa real digraphs do not agree with the adapted principle of consistency mooted by this study. Furthermore, what some scholars consider as digraphs are in the real sense consonant clusters thus their categorization as digraphs flout the basic description of digraph. As a result, this study advocates for a deliberate harmonization of Hausa orthography in such a way that orthographic reform could be engendered in respect of the use of digraph. In consequence, the following are recommended for adoption or adaptation for corpus planning:

That Hausa orthography should drop the use of digraphs so as to check the confusion and inconsistencies the status quo has generated;

That the digraph sh should be replaced with a monograph or an apostrophe or diacritic carrying letter (e.g. s', š or ʃ) so as to engender consistency with the other letters that use one symbol to represent one sound. Note that Hausa already has hooked and apostrophe carrying consonants.

That the digraph ts should be replaced with a monograph or an apostrophe or diacritic carrying letter (e.g. t', t, or ɬ) so as to engender consistency with the other letters that use one symbol to represent one sound. Note that the phonological alphabetic representation of ts as presented by Amfani (2007) is t'.

That the Hausa consonant clusters should be dropped from the list of Hausa letters. Teachers of Hausa and applied Hausa linguists should treat them under
the phonotactics of Hausa consonants. Note that English language does not present consonant clusters and digraph in the list of its letters (scholars are calling for the dropping of q and x from the list of English letters). In addition, consonant clusters are not listed in the list of letters of Hausa ajami.

That applied Hausa linguists should work out computer/typewriter keys that will take care of the proposed orthographic reforms in respect of the real digraphs. Professor Abdallah Uba Adamu has done this for the Hausa hooked letters. His expertise should be sought in respect of the present effort.

References


